

# A Preliminary Study on the Use of Vocal Function Exercises to Improve Voice in Male-to-Female Transgender Clients

\*Marylou Pausewang Gelfer and †Bethany Ramsey Van Dong, \*Milwaukee, †Wausau, Wisconsin

**Summary: Objectives.** This study explores the outcomes of symptomatic voice treatment plus Stemple's vocal function exercises (VFEs) for a group of male-to-female (MTF) transgender (TG) clients seeking voice feminization. Both acoustic and perceptual outcomes were assessed, in addition to the clients' attitudes toward VFE.

**Design.** Prospective treatment study.

**Method.** Three MTF TG clients plus three control female speakers and three control male speakers served as subjects. All provided a variety of speech samples. The TG clients underwent symptomatic voice therapy for 6 weeks, while simultaneously performing the VFE protocol. At the end of therapy, the TG clients provided posttreatment voice samples. All voice samples were analyzed for speaking fundamental frequency (SFF), SFF upper and lower limits, and the first three formants of /i/. A CD of pre- and posttreatment voice samples plus the control voices was presented to listeners for gender judgments and masculinity and femininity ratings.

**Results.** For acoustic measures, the TG subjects appeared more similar to the male control speakers in the pretest, and more similar to the female controls in the posttest. Perceptually, listeners continued to identify the TG subjects as male following therapy, although they were rated as significantly less masculine and more feminine. TG subjects were generally positive about the addition of VFE to their therapy experience.

**Conclusions.** The addition of VFE did not appear to improve posttreatment outcomes compared with previous literature. It was suggested that both number of sessions and experience living full-time as a woman might be important variables in predicting progress in therapy.

**Key Words:** Transgender voice—Transsexual voice—Voice therapy outcomes—Vocal function exercises.

## INTRODUCTION

The term “transgender” (TG) is a broad one, encompassing anyone who strays beyond the societal norm of binary gender.<sup>1</sup> Included under the umbrella term “TG” are those individuals who believe that their biological sex does not match their psychological orientation and who take steps to reconcile this difference by physically transitioning to their psychological gender. Most who do so are males wanting to be reassigned as females.<sup>2,3</sup> The transition process is a complex one, and not all male-to-female (MTF) TG individuals choose this path. However, those who do are likely to seek out intervention services such as hormone treatments, electrolysis, cosmetic surgery, make-up and clothing coaching, and voice and communication therapy,<sup>4</sup> as well as surgical sexual reassignment.

The speech-language pathologist is often called on during the transition period for MTF TG persons. For those transitioning in the other direction, from female to male (FTM), hormone treatment has an effect on the larynx that ultimately lowers pitch to a more masculine range, so this group is less likely to seek services for vocal intervention.<sup>5</sup> However, because hormone treatment for MTF persons does not affect vocal pitch, voice therapy is often sought to develop a more socially acceptable “feminine” voice.<sup>6</sup>

Research on voice therapy outcomes for MTF TG individuals has shown varying degrees of success in voice feminization. A growing number of studies have examined the acoustic outcomes of voice therapy for MTF TG individuals; but these investigations have typically focused more on outcomes and less on the effects of a specific therapy type. For example, Dacakis<sup>7</sup> showed that for 10 MTF TG clients, speaking fundamental frequency (SFF) was 125.5 Hz at the start of therapy, 168.1 Hz at discharge (after 10–90 sessions of therapy), and 148.6 Hz at follow-up (1–8.9 years postdischarge, with an average of 4.3 years postdischarge). This research, the first of its kind on long-term outcomes for MTF clients, nevertheless provided very little description on the therapy used, saying only that it focused “primarily on increasing mean fundamental frequency” (p. 551).

Soderpalm et al<sup>8</sup> were somewhat more specific about the therapy used in their long-term study. According to Soderpalm et al,<sup>8</sup> each therapy session lasted for 45–60 minutes and comprised two parts: vocal hygiene exercises, including relaxation, breathing, and phonation balanced in expiratory and laryngeal muscular effort; and phonatory/articulatory exercises, including gradual pitch climbing, improving articulatory contacts, and encouraging anterior articulatory placement. It was noted that therapy was loosely based on the accent method. This was an extensive study involving 25 subjects, but for comparison purposes, the present authors extracted a subset of nine participants who were MTF TG individuals, received voice therapy as part of the study (not all subjects in the study received therapy), and had baseline, intervention, and follow-up testing. For these nine participants, a pretest SFF of 138.8 was found. After therapy (1–49 months, mean = 12.1 months), SFF rose to

Accepted for publication July 12, 2012.

From the \*Department of Communication Sciences and Disorders, University of Wisconsin—Milwaukee, Milwaukee, WI; and the †Wausau School District, Wausau, WI.

Address correspondence and reprint requests to Marylou Pausewang Gelfer, Department of Communication Sciences and Disorders, University of Wisconsin—Milwaukee, PO Box 413, Milwaukee, WI 53211. E-mail: gelfer@uwm.edu

Journal of Voice, Vol. 27, No. 3, pp. 321–334

0892-1997/\$36.00

© 2013 The Voice Foundation

<http://dx.doi.org/10.1016/j.jvoice.2012.07.008>

148.2 Hz and rose again in the follow-up period (14 months–6 years postintervention, mean = 27.4 months) to 157.3 Hz.

The studies of Dacakis<sup>7</sup> and Soderpalm et al<sup>8</sup> were milestones in that they provided acoustic data on short- and long-term outcomes of voice therapy for MTF TG individuals. However, neither study provided perceptual data to assess listener responses to the voice changes made by subjects. This is important because acoustic measures alone, although important, do not provide adequate information on what clients are most concerned about: the reactions of listeners to their voices. In addition, neither study emphasized the specific type of therapy provided to the subjects. Because both studies were retrospective, using existing clinical records that had come from treatment sessions over a period of years, it may be that therapy approach was not well controlled enough for this to have been a factor in the research design.

A few recent prospective studies have examined listener perceptual judgments, with a specific therapy approach incorporated into the study design. One of these studies, Gelfer and Tice,<sup>9</sup> included five MTF TG clients who participated in an average of 15.4 therapy sessions (range = 13–16 sessions). The therapy administered adhered to an approach outlined by Gelfer,<sup>10</sup> which emphasized raising pitch, use of a light and clear voice quality, and use of feminine intonation patterns at various loudness levels, while moving from /m/-initiated syllables to phonemically unrestricted words, phrases, and sentences, to multisentence utterances. In Gelfer and Tice,<sup>9</sup> both perceptual judgments and acoustic measures were used to compare pretest voice samples with samples taken immediately after therapy (the immediate posttest) and voice samples gathered 15 months after the termination of therapy (the long-term posttest).

Perceptual results from Gelfer and Tice<sup>9</sup> revealed that in the pretest, audio voice samples from MTF TG clients were perceived by listeners as produced by female speakers by an average of 1.9% of the time. In the immediate posttest, 50.8% of the samples were perceived as being produced by female speakers. In the long-term posttest, female perception fell to 33.1%. Ratings of the subjects on scales of masculinity (1 = very masculine and 7 = not at all masculine) and femininity (1 = very feminine and 7 = not at all feminine) showed that the TG speakers were rated as sounding less masculine and more feminine to a significant degree both in the immediate posttest and in the long-term posttest compared with the pretest (although scores were not quite as favorable in the long-term posttest). Acoustic measures of fundamental frequency and its variability as well as vowel formant frequency values of the TG clients were generally similar to those of male control subjects in the pretest, similar to female control subjects in the immediate posttest, and somewhere in between at the long-term posttest. However, there were considerable differences among subjects, with some participants showing marked gains in voice feminization, whereas others showed limited gains. These results suggested that the therapy techniques described by Gelfer<sup>10</sup> could result in positive change in the voices of MTF TG clients; but it was also clear that there was much potential for improvement of outcomes in terms of degree of progress and consistency across subjects.

A different approach to voice therapy for MTF TG individuals was taken by Carew et al.<sup>11</sup> These researchers emphasized oral resonance with a focus on raising vowel formant frequencies. Because this was prospective research, therapy approach was incorporated into the study design. Subjects included 10 MTF TG individuals, each of whom participated in five therapy sessions that targeted lip spreading and forward tongue carriage. These researchers did not examine listeners' perceptions of speaker gender (male vs female), but they did include listener judgments of masculinity-femininity on a single scale (where a rating of 0 = very masculine and a rating of 10 = very feminine). Their results showed that four subjects were consistently rated as more feminine in the posttest compared with the pretest, three were rated as more feminine inconsistently, and three were rated either the same in the pretest as in the posttest or more masculine. As in the research of Gelfer and Tice,<sup>9</sup> individual differences based on speaker were evident in both the absolute values of listener' ratings of masculinity-femininity, and the degree of difference in ratings between pre- and posttests. Acoustic measures revealed that mean measures of vowel formant frequencies increased significantly from pre- to posttest for F1 for the vowels /a/ and /u/, for F2 for the vowel /a/, and for F3 for all vowels (/i/, /a/, and /u/).

It was further noted that SFF rose from 119.4 Hz in the pretest to 133.3 Hz in the posttest, despite not being directly addressed in therapy.

Although Gelfer and Tice<sup>9</sup> and Carew et al<sup>11</sup> used different therapy techniques, both studies adhered to what Andrews<sup>12</sup> has called the "symptomatic voice treatment approach." In this type of voice treatment, overt vocal behavioral characteristics are directly modified using facilitating techniques designed to elicit the desired vocal behaviors. The first step in this type of treatment is to identify the vocal behaviors that need to be modified. For MTF TG individuals, one vocal behavior typically targeted in therapy has been vocal pitch. This selection is based on previous research,<sup>13–17</sup> which showed that raising pitch was important to perception of an MTF TG speaker as female. In addition, vowel formant frequencies are also frequently targeted, based on research which has showed that TG individuals perceived as female typically have higher vowel formant frequencies than those perceived inconsistently or as male (eg, Refs<sup>15,18,19</sup>). Other variables, such as upper and lower limits of frequency and intonation patterns,<sup>14,15</sup> and the vocal quality of breathiness<sup>20</sup> have been investigated in terms of their effect on listener judgments of the masculinity and femininity of voice samples, and are occasionally addressed in voice therapy as well. By focusing on specific vocal parameters to be changed, both Gelfer and Tice<sup>9</sup> and Carew et al<sup>11</sup> used what could be classified as symptomatic voice therapy approaches, although each study emphasized a different primary aspect of voice.

Another approach to voice therapy describe by Andrews<sup>12</sup> is the physiological approach. This type of therapy is based on the idea that if the voice production mechanism is balanced in terms of the activity of various muscle groups, vocal health and the voice will improve. This approach uses physical exercises and manipulations,<sup>12</sup> and unlike the symptomatic

approach, does not target particular vocal symptoms. The principles of exercise physiology, such as warm-ups, exercise of various intensities, rest periods, and cool-downs, are also sometimes included to balance and condition the voice production muscles.

One example of a physiological approach to voice therapy is Stemple's vocal function exercises (VFEs).<sup>21</sup> Voice clients who are taught VFE learn a series of four vocal exercises, which are to be done two times each, twice per day. These exercises theoretically improve the strength, endurance, and coordination of the respiratory, phonatory, and resonance systems, and improve maximum phonation time, glottic closure, and phonatory efficiency.<sup>22,23</sup> VFE have been found to improve aerodynamic function in elderly men,<sup>24</sup> self-perception of handicap and listener's judgments of dysphonia in elderly men and women,<sup>22</sup> and acoustic and perceptual measures of the voices of Vietnamese elementary school teachers with muscle tension dysphonia.<sup>25</sup>

Research evidence that VFE can improve the voices of those with dysphonia is of interest because the vocal changes attempted by MTF TG individuals to feminize their voices may lead to vocal fatigue and other physical complications. For example, Soderpalm et al<sup>8</sup> found "moderate-to-pronounced supraglottal constriction" (p. 25) in about half of their total population of 25 TG clients/subjects during pretherapy laryngeal examinations. McNeill et al<sup>26</sup> found that on the Voice Handicap Index, a self-evaluation instrument for assessing the impact of a voice disorder, mean scores across all subjects fell within the range of mild vocal dysfunction. However, on the "Physical" subscale, which included rating items such as "My voice feels creaky and dry," and "I run out of air when I talk," McNeill et al<sup>26</sup> got more negative results, with some individual subjects' scores falling into the severe range.

For the MTF TG population, increased respiratory support and efficiency of laryngeal valving may help maintain a smooth voice quality while adjustments to vocal pitch are accomplished. These physiological elements are sometimes addressed in symptom-based voice therapy, but it is possible that the expanded emphasis on strengthening, balancing, and coordinating the respiratory, phonatory, and resonance systems provided by VFEs may help TG individuals more quickly acquire feminine vocal behaviors.

The purpose of this study was to determine the benefits of VFEs used concurrently with symptomatic voice therapy (as described in Gelfer<sup>10</sup>) on acoustic and perceptual measures of voice for TG individuals. Specific research questions included the following:

1. What acoustic outcomes are achieved when MTF TG individuals are provided with a combination of symptomatic voice treatment and VFE? Will the addition of the VFE protocol result in improved acoustic posttreatment outcomes for MTF TG individuals compared with previous literature?
2. What perceptual results occur when MTF TG individuals are provided with a combination of symptomatic voice treatment and VFE? Will the addition of the VFE proto-

col result in improved perceptual posttreatment outcomes for MTF TG individuals compared with previous literature?

3. Do TG clients believe that the VFE protocol was helpful to them throughout the therapy process?

The results of research of this type, although preliminary in nature, may eventually impact how speech-language clinicians approach voice therapy with MTF TG individuals to ensure optimal benefit for the client.

## METHOD

### Participants

**TG participants.** The speaker subjects recruited for this study included three MTF TG individuals who met the following inclusion criteria: self-identified as TG, had counseling for gender change before the start of the study, currently living as a woman full-time or undergoing hormone therapy with plans to live full-time as a woman within the next year, native speaker of American English, bilateral hearing within normal limits, speech and voice characteristics within normal limits for a male speaker, perceptually identified as male speakers by the investigators, and a nonsmoker. In addition, subjects were excluded from the study if they reported a history of previous voice therapy for voice feminization, phonosurgery, voice disorders, neurological disorders, or cancer of the head/neck. Mean age for the TG subjects in this study was 43 (years):1 (months), with a range of 32:11–50:5. Mean height averaged 5'10", with a range of 5'8"–6'0". Of the three subjects, one had been living as a woman for approximately 7 months at the start of the study, one had just begun living as a woman full-time the same month as the study started, and one had had both counseling and hormone treatments for the past 10 months and planned to initiate going "full-time" the next year. The TG participants participated in 6 weeks of therapy (12 sessions) and provided voice samples pre- and posttherapy for acoustic and perceptual analysis.

**Control participants.** Three non-TG males and three non-TG females were recruited as control speakers. Inclusionary criteria for control subjects were: native speaker of American English, bilateral hearing within normal limits, speech and voice characteristics within normal limits for their gender, and nonsmoker. Control subjects were excluded if they reported a history of voice or neurological disorders or head/neck cancer. Additional inclusion criteria for the control participants were that they had to match one of the TG speakers in both height (within 2 in.) and age (within 6 years). Thus, each TG speaker had one male control speaker and one female control speaker who matched her in terms of height and age. At the conclusion of the selection process, mean age of the control female group was 42:4 (range = 38:1–44:11), and mean height was 5'8.5" (range = 5'7"–5'10"). Mean age of the control male group was 43:11 (range = 38:9–46:8), and mean height was 5'10" (range = 5'7"–6'1"). Control speakers did not receive any type of intervention services and were seen only once to record voice samples that were later subjected to perceptual and acoustic analysis.

**Listener participants.** Listeners consisted of 27 college students who were recruited from large health sciences, psychology, statistics, and other classes. All listener participants met the following criteria: normal hearing in both ears, native speaker of American English, no previous coursework in the field of communication sciences and disorders, aged between 18 and 35 years, and met all reliability criteria related to male-female identification and masculinity and femininity scale correlations (see Results section). This group ultimately included 13 males with an average age of 21:5 (range = 18:2–31:10) and 14 females with an average age of 21:5 (range = 18:7–31:2). Listener subjects rated the TG voices from the pre- and posttest, as well as control male and female voices.

### Voice data collection procedures

The TG speakers were initially screened over the telephone for inclusion and exclusion criteria. Those who initially met the criteria were asked to meet with the researchers for an explanation of the study, and if agreeable to participation, for the signing of informed consent. Hearing<sup>27</sup> and voice/speech screenings using the Consensus Auditory-Perceptual Evaluation of Voice<sup>28</sup> were then performed, and if all the criteria were met, the individual was enrolled in the study.

To provide pretreatment voice samples, TG speaker subjects were asked to read the Rainbow Passage,<sup>29</sup> provide a 30-second spontaneous speech sample, and read 10 semispontaneous question/answer (Q/A) sets. These speech samples were collected with participants seated in a quiet acoustically treated room, in front of a Shure microphone (Model SM58; Shure Inc., Niles, IL) placed 10 in. from the mouth. An AudioBuddy Dual Mic PreAmp Direct Box (Midiman U.S., Arcadia, CA) preamplifier connected to the microphone fed into a Dell Optiplex GX280 computer. The *Real-Time Pitch* application (Model 5121, version 3.1.6, 2000–2006; Kay PENTAX, Montvale, NJ) of the *Multi-Speech* acoustic analysis program (Model 3700, 3.1.6, 2000–2006; Kay PENTAX) was used to record and save the speech samples. These samples were later retrieved and analyzed for SFF, SFF range, upper and lower limits of SFF, pitch sigma (individual standard deviation of SFF), and the first three formants of /i/ from the word “beach” in the selected semispontaneous Q/A set. The same tasks and measures were repeated during individual sessions held at the termination of therapy to obtain posttherapy measures and determine change over time.

Male and female control participants were also screened over the phone or in person to see if they initially met the criteria of the study, and if they appeared to be a match for one of the TG subjects. If they did, the study was explained to them, and they were asked to sign informed consent. The hearing and voice/speech screening protocols were then administered, and if participants met all of the inclusion criteria and none of the exclusion criteria for the study, they were asked to perform the same speaking tasks as the TG participants: reading the Rainbow Passage, providing a 30-second spontaneous speech sample, and producing 10 semispontaneous Q/A sets. Their samples were recorded and saved using the procedures outlined above for TG subjects. The control subjects were not required to re-

turn a second time as they did not participate in any type of therapy program and thus there were no posttreatment samples to gather.

### Therapy procedures for TG participants

The TG participants in this study received individual voice therapy for two 1-hour sessions per week for 6 weeks (for a total of 12 hours of therapy) and were also required to do Stemple's VFEs two times each, twice per day at home, for the entire 6-week experimental period. The voice therapy approach used was symptomatic voice therapy,<sup>12</sup> based on the model outlined in Gelfer.<sup>10</sup> An initial target SFF was chosen for each participant based on age, vocal range, and initial SFF measurements. After the target was established, the subject began by chanting syllables beginning with /m/, /n/, and /l/. After the syllable level was mastered, the subject moved on to chanting words with the same phonemic restrictions, producing words with speech intonation, putting words together into phrases, producing sentences, producing sentences with varying emotionality, producing sentences with unrestricted phonemic contexts, reading paragraphs, and generating spontaneous speech. The *Real-Time Pitch* program of *Multi-Speech* was used in each session to provide immediate feedback to both the clinician and subject. In addition to this therapy, one training session and a second follow-up session concerning implementation of the home program of VFEs<sup>30</sup> were conducted.

### Description of VFE protocol

The VFEs protocol<sup>30</sup> consisted of the following:

1. On the musical note middle C (262 Hz, semitone [ST] #48<sup>31</sup> participants were instructed to hold the vowel /i/ as softly as they could, for as long as possible. The goal was to maintain extreme forward focus without nasality. This was considered a warm-up exercise for the intrinsic laryngeal muscles. Each time the exercise was completed, subjects were required to record total duration of the prolongation in seconds.
2. On the word “knoll,” participants were to glide from their lowest note possible to their highest note. The participants were told that the goal of this exercise was to perform the sliding scale with no voice breaks. If a break did occur, the participant was told to continue the glide without stopping. If the voice break occurred at the top of the participant's current range, they were told to continue the exercise without voice because, according to Stemple,<sup>30</sup> the vocal folds would continue to stretch. Again, an extreme forward focus was encouraged. This exercise was meant to stretch all intrinsic laryngeal muscles with a focus on the gradual engagement of the cricothyroid muscle. Again, subjects were required to record the duration of each glide in total number of seconds.
3. On the word “knoll,” participants were to glide from their highest note to their lowest possible note. Participants were told to perform the exercise with the goal of no voice breaks and with an extreme forward focus. As with the previous exercise, the participants were told to

continue if a break did occur. This downward glide exercise was meant to encourage a gradual engagement of the external thyroarytenoid muscle. Subjects were required to record the total number of seconds of duration for each glide.

4. Beginning on the G below middle C (196 Hz, ST #43), each participant was told to hold each note (G, A, B, C, and D) for as long as possible on the vowel /o/. The goal was the same as in the first exercise, to maintain the note for as long as possible, as softly as possible, and with an extreme forward focus.

Participants were instructed to do all exercises using as quiet a voice as possible while still producing voicing (not a whisper). Each individual exercise was done two times before moving on to the next one. The entire protocol was done twice daily, once in the morning and once at night. Home logs were provided to subjects to keep a record of VFE compliance and weekly phonation durations. Participants were asked to return the home log on their first day of therapy each week.

Written instructions of the VFE procedure were provided to the participants. The initial training consisted of the first author demonstrating each exercise and explaining proper implementation. A CD with verbal instructions and pitch exemplars was also given to each participant for home use. A follow-up retraining session was done at the third week of therapy. At this time, each participant performed the exercises for the authors to ensure that procedures were being implemented correctly.

### Client evaluation of VFEs

At the same time that the posttherapy voice data collection occurred, all TG participants were asked to fill out a questionnaire regarding their opinions on the effectiveness of the VFE home protocol, its ease of implementation, and overall impressions of therapy. Each question was answered on a scale of 1–5 (1 = not at all and 5 = very much). Space was provided for comments after each question.

### Acoustic analysis

The speech samples of the TG subjects (pre- and posttest) and control participants were subjected to acoustic analysis at the end of the therapy period. All samples were analyzed using the *Real-Time Pitch* application of the *Multi-Speech* acoustic analysis program. Measures of SFF, upper limit of SFF, lower limit of SFF, and pitch sigma were recorded for each spontaneous speech sample, reading of the Rainbow Passage, and the semispontaneous Q/A sets. The SFF, upper limit of SFF, and lower limit of SFF were selected as dependent variables in this study because they have been shown in previous studies to be a reliable measure of comparison between pre- and posttest data. Pitch sigma, or pitch standard deviation, was chosen because it is an estimate of SFF variability that is not as sensitive to artifacts (nonvoiced high and low frequencies) as upper and lower limits of SFF, and thus may be a more valid index of vocal variability.

The *Multi-Speech* acoustic analysis program was used to obtain the first, second, and third vowel formants of the vowel /i/

extracted from the first occurrence of the word “beach” from one of the semispontaneous Q/A samples (“Do you find shells at the park or at the beach? I find shells at the beach”). Each vowel segment was downsampled to a rate of 11 kHz. The investigators created a time-by-frequency spectrogram of the most stable middle portion of each vowel with the initial and final consonants removed, usually 80–100 milliseconds in duration. A long-term average spectrum (LTAS) analysis for the sample region was derived, and a linear predictive coding (LPC) analysis was calculated by the program to identify the first three vowel formants. These values were compared with normative vowel formant data for males and females as presented by Hillenbrand et al.<sup>32</sup> In addition to the formants identified by the program from the LPC analysis, the investigators used two additional methods to determine vowel formant frequency values: they independently identified peaks corresponding to formant center frequencies from the LTAS; and they identified formant frequencies via cursor from three points (near the beginning, the middle, and the end of the vowel) in the spectrogram. The formant data from these points in the spectrogram were averaged, and the mean was then averaged with the investigator-identified values from the LTAS analysis and the program-identified values from the LPC analysis. This procedure was determined necessary by the authors to obtain reliable identification of vowel formant frequencies owing to the brief nature and variability of vowels extracted from running speech.

### Perceptual protocol

**Construction of the stimulus CD.** Speakers’ productions of a short semispontaneous Q/A set were the stimuli recorded on a CD for listeners’ perceptual judgments. The same Q/A set was used for each speaker (set #9: “Do you find shells at the beach or at the park? I find shells at the beach.”). This Q/A set was approximately 5–7 seconds in length for all speakers. Q/A sets for a particular speaker were repeated four consecutive times, with 3 seconds in between the subsequent three sets, and 5 seconds between the last playing of one speaker and the first playing of the next.

The CD heard by listener subjects contained a total of 24 Q/A sets: three pretest samples from the TG subjects, three posttest samples from the TG subjects, three samples from control males, and three samples from control females, each presented twice for reliability purposes. The entire stimulus set was presented in quasi-random order, that is, the order was determined by a random numbers table with the stipulation that identical samples could not occur together (eg, the first and second presentations of a particular TG speaker’s posttest sample).

**Listening procedure.** Listeners were seated in a quiet room in small groups of one to eight, with each individual approximately 60 in. from a speaker. The CD containing the experimental stimuli was presented via a Dell Latitude D600 Laptop computer and a Dell Zylux Multimedia Speaker System. Ambient noise level in the room was approximately 50 dB, and stimuli were presented at a comfortable listening level (approximately 70 dB at the ear). Listeners were asked

to provide judgments of the following: the gender of the speaker (male or female), the age of the speaker, rating on a scale of masculinity, rating on a scale of femininity, and rating on a scale of pleasantness.

Each rating scale judgment was on a seven-point equal-appearing interval scale, where one corresponded to very masculine for the first scale, very feminine for the second scale, and very pleasant for the third scale. A score of seven corresponded to not at all masculine, not at all feminine, or not at all pleasant for each scale, respectively. Listeners were instructed that a “very masculine” (or “very feminine” or “very pleasant”) voice should be rated with a score of one; and a voice that was “not at all masculine” (or “not at all feminine” or “not at all pleasant”) should be rated with a score of seven. They were further instructed that the voices that fell in between those extremes should be rated as the listener saw fit by circling the appropriate number between one and seven; and that listeners should use their own judgment and not consult with anyone else in the group. The age and pleasantness judgments were intended to be foils to distract the listeners from the investigators’ primary interest in gender identification and masculinity/femininity judgments.

## RESULTS

### Listener reliability

Listener reliability was measured in two ways: concordance of gender identification for pairs of TG participants’ voices, and correlations between first and second ratings of each voice sample. With respect to the first measure, listeners’ gender identifications of the 12 TG voice samples (three subjects in both the pre- and posttest, each presented twice) were organized into pairs of first presentation versus second presentation of each sample. Only listeners who were at least 83% concordant for gender identification were retained in the study, that is, only listeners who rated both presentations of a TG voice (eg, the first and second presentations of the pretest sample of subject 1) as the same gender, for five of the six TG voice pairs, were included in further analyses of listener judgments. This procedure was intended to remove listeners who did not appear to have stable internal criteria for “male” and “female” voices, and who seemed to be guessing at gender for the TG speakers.

Results of this procedure revealed that 22 of the selected 27 listeners were 100% concordant, or reliable, for gender identification, and five listeners were 83% reliable. Average concordance (reliability) for gender identification for TG voice samples was 97%. Gender identifications for the control male and female speakers were not included in this reliability calculation as no listener misidentified or was inconsistent in the identification of the gender of a control male or control female speaker.

With respect to the second measure, listeners’ ratings of all voice samples on the masculinity and femininity scales were examined for reliability using Pearson correlations.<sup>33</sup> The rating scale judgments made by the listeners were on seven-point equal-appearing interval scales and were considered by the investigators to be at the interval level of measurement.

The data sets for masculinity and femininity ratings were tested and found to meet the additional assumptions for parametric statistics.<sup>34</sup> Thus, parametric statistics were used for determining reliability. Criteria for inclusion in the study were that a listener’s rating scale judgments on the first presentation of all the stimulus voices had to correlate at  $r > 0.5$  ( $P < 0.05$ ) with their ratings on the second presentation. These criteria were applied to the masculinity rating scale and the femininity rating scale separately. Listeners had to be reliable on both scales to be retained in the study. Male and female control voices were included in this reliability procedure because a preliminary analysis revealed that variability between ratings of the first and second presentations of male and female control speakers on the masculinity and femininity scales was similar to the variability for TG speakers.

Pearson correlation results revealed that for the masculinity scale, coefficients for the selected listeners ranged from  $r = 0.664$  ( $P = 0.018$ ) to  $r = 0.958$  ( $P = 0.000$ ). For the femininity scale, correlation coefficients ranged from  $r = 0.699$  ( $P = 0.011$ ) to  $r = 0.994$  ( $P = 0.000$ ). These reliability results suggested that the 27 selected listeners were able to judge the gender of the TG voices consistently, with adequate internal criteria for “male” and “female” voices; and that they were reliable in their ratings of all voices on the masculinity and femininity scales.

### Acoustic outcomes

Acoustic outcomes for this study can be seen in Table 1. Visual inspection of the data reveals a strong similarity between the acoustic measures of the male control subjects and the pretest voices of the TG subjects. This is not surprising because one of the inclusion criteria for TG subjects was to have a perceptually male-sounding voice before the initiation of voice treatment. It can also be noted that posttest acoustic measures for the TG clients are more similar to the female control speakers than they are to the male control speakers. With only three subjects, it was not possible to do inferential statistics (parametric or nonparametric) to determine if significant differences occurred between the pre- and the posttest; however, it is clear that the posttest voice samples increased markedly compared with the pretest measures in terms of SFF.

### Perceptual outcomes

Several perceptual rating procedures were completed to determine the listeners’ response to the changes transgender subjects made in their voices from pre- to the posttest.

**Gender identification results for TG speakers.** When asked to judge the speakers’ gender, the 27 selected listeners judged all three TG speakers to be “male” in the pretest. Specifically, each of the 27 listeners judged each voice twice, for a total of 162 judgments ( $27 \times 3 \times 2$ ). All 162 judgments of the pretest voices identified the speaker as male. In the posttest, there were a total of 150 identifications of the speakers as male and 12 identifications as female, for a total of 92.6% male and 7.4% female. Within the three speakers, there was a range of female identifications in the posttest: one subject was identified as

**TABLE 1.****Acoustic Measures for Control Male Speaker Subjects (N = 3), Pretest Voices of Transgender Speakers, Posttest Voices of Transgender Speakers (N = 3), and Control Female Speakers (N = 3)**

Acoustic Measures	Control Male Speakers	Transgender Speakers, Pretest	Transgender Speakers, Posttest	Control Female Speakers
Spontaneous speech				
SFF (Hz)	110.79	115.53	152.83	179.10
SFF upper limit (Hz)	223.42	225.83	386.51	261.13
SFF lower limit (Hz)	93.58	99.97	113.39	125.24
Pitch sigma (ST)	1.9	3.2	2.2	3.1
Rainbow passage				
SFF (Hz)	115.44	122.04	177.09	175.54
SFF upper limit (Hz)	215.11	240.41	388.92	447.74
SFF lower limit (Hz)	90.30	92.79	132.98	121.02
Pitch sigma (ST)	2.2	3.0	2.3	2.7
Semispontaneous Q/A sets				
SFF (Hz)	122.86	124.52	183.02	199.70
SFF upper limit (Hz)	152.15	205.86	298.84	387.48
SFF lower limit (Hz)	100.30	93.90	145.29	142.20
Pitch Sigma (ST)	2.1	2.8	2.3	3.5
F1 of /i/ (Hz)	294.22	298.99	353.01	386.10
F2 of /i/ (Hz)	2170.95	2188.79	2322.73	2663.77
F3 of /i/ (Hz)	2869.28	2640.09	2987.42	3092.44

being a female 14.8% of the time (8/54), one subject was identified as female 7.4% of the time (4/54), and one was consistently identified as male.

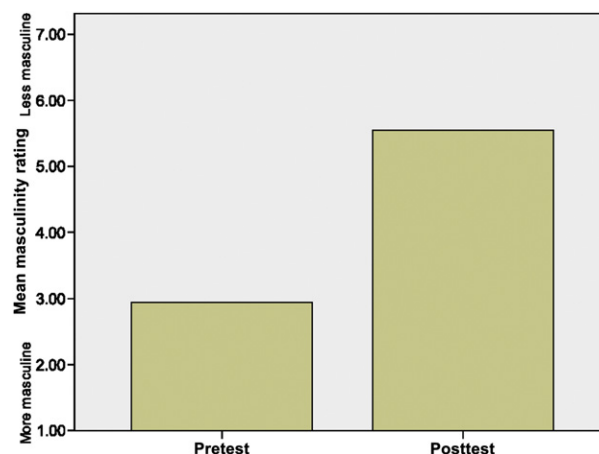
**Masculinity and femininity rating scale results.** Listeners also rated each voice on two rating scales related to gender: first was a seven-point masculinity scale, where 1 = very masculine and 7 = not at all masculine; and second on a seven-point femininity scale (1 = very feminine and 7 = not at all feminine). For the TG speakers, masculinity and femininity ratings were made in both the pre- and the posttest. The male and female control voices were also rated.

For these analyses, listeners' ratings of the first and second presentation of each voice sample were averaged to create a single value for each listener for each voice. In addition, as with the reliability statistics, the rating scale results were considered to be interval-level data (based on equal-appearing interval scales); and testing revealed that both sets of listeners' ratings for both scales met the assumptions for parametric statistics<sup>34</sup>, so parametric statistics were used both descriptively and inferentially for the rating scale data.<sup>33</sup>

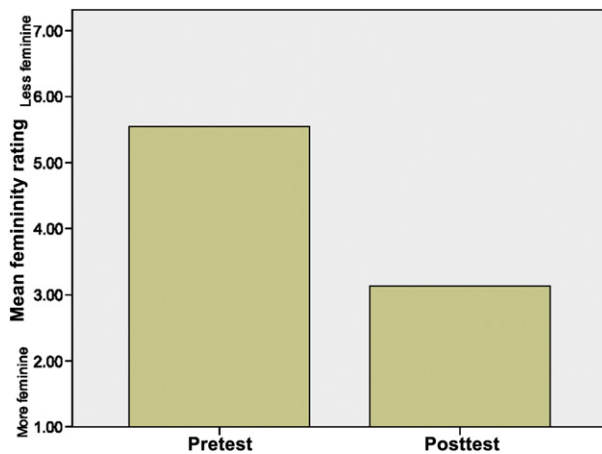
The results of the listeners' rating scale judgments for the masculinity and femininity scales can be seen in Figures 1 and 2. The mean of judgments on the masculinity rating scale based on 27 listeners and three TG pretest voice samples was 2.94 on a scale of one to seven, where 1 = "very masculine" and 7 = "not at all masculine." In contrast, in the posttest, the mean masculinity rating for the three TG samples was closer to the "not at all masculine" end of the scale at 5.55 (Figure 1). For the femininity rating scale, the pretest mean was 5.48, closer to the "not at all feminine" end of the rating scale. In the posttest, TG voices were judged as closer to the "very feminine" end of the scale with a mean of 3.07 (Figure 2).

To test the statistical significance of the pre- to posttest changes seen in the masculinity and femininity rating scale data, 2 two-way analyses of variance (ANOVA) were performed, with speaker as a between-subjects independent variable, therapy status (pre- and posttest) as a within-subjects independent variable, and speaker  $\times$  therapy status as the interaction term.<sup>33</sup> A level of  $P < 0.05$  was selected as the threshold for significance.

In the first ANOVA, listeners' ratings of the TG voice samples of the pre- versus posttest on the masculinity rating scale were compared. Results can be seen in Table 2. In this analysis, main effects for therapy status and subject were significant, as well as the therapy status  $\times$  subject interaction. This result



**FIGURE 1.** Comparison of mean masculinity ratings for the pre- and posttest as provided by listeners (N = 27). On this scale, 1 = very masculine and 7 = not at all masculine.



**FIGURE 2.** Comparison of mean femininity ratings for the pre- and posttest as provided by listeners ( $N = 27$ ). On this scale, 1 = very feminine and 7 = not at all feminine.

indicated that TG voices were perceived as significantly less masculine in the posttest compared with the pretest. The significant interaction indicates that individual TG speakers changed to significantly different degrees in terms of the perceived masculinity of their voices.

In the second ANOVA, listeners' ratings for the TG voice samples of the pre- versus posttest on the femininity rating scale were compared. Results can be seen in Table 3. Results again revealed significant main effects for therapy status and subject, as well as a significant therapy status  $\times$  subject interaction. The analysis indicated that TG voices were perceived as significantly more feminine in the posttest compared with the pretest, with some individual TG speakers making more progress than others in increasing the femininity of their voice.

### Outcomes of VFEs questionnaire

The Questionnaire on Voice Function Exercises and the combined results of TG subjects' responses are presented in the Appendix section. From the subjects' comments, it is clear that they felt VFE was a positive part of the therapy experience, but that they believed the exercises alone would not have resulted in the same level of voice feminization as the combination of individual symptomatic voice therapy plus VFE.

## DISCUSSION

In this study, three TG individuals underwent symptomatic voice therapy for a period of 6 weeks (12 sessions) while at the same time performing Stemple's VFEs twice each day. Pre- and posttest acoustic and perceptual measures were compared to assess subjects' gains during therapy and to gather preliminary data regarding the effects of inclusion of a physiological type of voice therapy in addition to symptomatic voice therapy on therapy outcome. When pre- and posttest results were compared, the TG subjects' acoustic measures clearly shifted, from being similar to control male voices in the pretest to being similar to control female voices in the posttest. In terms of perceptual changes, the outcome was somewhat mixed. None of the TG speakers was perceived by listeners as being female in the pretest; however, in the posttest, only 7.4% of the voices were perceived as female. On the other hand, all TG speakers were rated by listeners as being significantly less masculine and more feminine in the posttest.

One research question for this study was if the addition of the VFE protocol to symptomatic voice therapy would result in improved acoustic posttreatment outcomes for MTF TG individuals compared with previous studies using symptomatic voice therapy alone. Although it was difficult in some cases to determine the specifics of the therapy approach in previous research, a comparison of acoustic outcomes from the literature is shown in Table 4. All the studies listed in Table 4 are similar in that all appeared to emphasize symptomatic voice therapy with the primary goal of raising SFF. As can be seen from the summarized results, the subjects in the present study did not raise SFF to the degree that subjects in Meszaros et al<sup>6</sup> and Gelfer and Tice<sup>9</sup> did. It is important to keep in mind that the number of subjects, number of sessions, variability in number of sessions, and specific therapy practices were very diverse among the cited studies. However, the addition of VFE to a symptomatic voice therapy protocol did not appear to have a markedly positive effect on raising SFF compared with other studies that did not use VFE.

The second research question of this study asked whether adding the VFE protocol to symptomatic voice therapy would result in improved perceptual posttreatment outcomes for MTF TG individuals compared with previous literature. In

**TABLE 2.**  
**Analysis of Variance for Pre- and Posttest Ratings on the Masculinity Scale**

Source	Type III Sum of Squares	DF	Mean Square	F	Significance
Tests of within-subjects effects					
Therapy status	274.821	1	274.821	472.468	0.000
Therapy status $\times$ subject	5.559	2	2.779	4.778	0.011
Error (therapy status)	45.370	78	0.582		
Tests of between-subjects effects					
Intercept	2921.877	1	2921.877	2079.578	0.000
Subject	42.281	2	21.140	15.046	0.000
Error	109.593	78	1.405		

Abbreviation: DF, degrees of freedom.

**TABLE 3.**  
**Analysis of Variance for Pre- and Posttest Ratings on the Femininity Scale**

Source	Type III Sum of Squares	DF	Mean Square	F	Significance
<b>Tests of within-subjects effects</b>					
Therapy status	234.722	1	234.722	301.786	0.000
Therapy status $\times$ subject	5.361	2	2.681	3.446	0.037
Error (therapy status)	60.667	78	0.778		
<b>Tests of between-subjects effects</b>					
Intercept	2964.500	1	2964.500	1294.069	0.000
Subject	35.065	2	17.532	7.653	0.001
Error	178.685	78	2.291		

Abbreviation: DF, degrees of freedom.

terms of gender identification, the only other study that has investigated this question was Gelfer and Tice.<sup>9</sup> In Gelfer and Tice's pretest, listeners judged the TG voices to be produced by female speakers 1.9% of the time. In the immediate posttest (the most comparable time interval to the present study), TG voices were identified as being produced by females 50.8% of the time. Those results were much more positive than the present study, where listeners perceived the TG voices as female 0.0% of the time in the pretest and 7.4% of the time in the posttest. However, in terms of masculinity and femininity judgments, the study of Gelfer and Tice<sup>9</sup> was very consistent with the present study. Both studies found significant differences between pre- and posttest listener's judgments, on a scale of masculinity and a scale of femininity. All TG speakers in both studies were perceived to be significantly more feminine and less masculine in the posttest compared with the pretest.

A third research question addressed the issue of TG clients' response to Stemple's VFEs. Results of client responses on a posttest questionnaire revealed a positive orientation to VFE. However, in the opinion of the clients, individual symptomatic therapy was more important to satisfactory progress in voice feminization.

Results of this preliminary study suggested that the addition of VFE to the more traditional symptomatic voice therapy focusing on pitch, vocal quality, and intonation did not markedly improve therapy outcome, when compared with the results of previous literature. Both acoustic and perceptual measures showed that the current TG subjects had all made progress toward voice feminization; but their progress was comparable with the progress reported in other studies where a VFE component was not used.

Consistent with previous investigations,<sup>7,9,11</sup> the present researchers found a considerable degree of variability in results among their speakers. In each study, it was clear that some individual TG subjects made more progress than others in raising SFF and its various components (and/or vowel formant frequencies), and in altering the perception of their voices from male to female. The variability in success rate among TG clients seen in Gelfer and Tice,<sup>9</sup> Carew et al,<sup>11</sup> and the present study is notable because all of those investigations were prospective in nature, and subjects within each study received the same number of treatment sessions (15, 5, and 12,

respectively). Thus, number of sessions was not a factor in individual outcome in any of these studies.

Other researchers have also noted the equivocal influence of number of sessions on outcome. For example, Dacakis<sup>7</sup> correlated actual SFF with number of treatment sessions and found a nonsignificant correlation of  $r = 0.474$ . She correlated maintenance of SFF and number of treatment sessions and found that there was a stronger correlation between number of treatment sessions and SFF maintenance ( $r = 0.745$ ,  $P < 0.05$ ); however, the correlation weakened considerably when one subject, an outlier who received 90 treatment sessions, was removed from the analysis ( $r = 0.476$ ,  $P < 0.05$ ). Similarly, Soderpalm et al<sup>8</sup> compared subjects in their treatment study who had had less than 14 sessions with those who had had more than 14 sessions and found a slight but nonsignificant increase in SFF for the over-14 sessions group compared with the under-14 sessions group. These results suggest that although the number of treatment sessions has some influence on voice treatment outcomes, there are other issues to consider.

Another important factor in degree of progress for individual TG clients in voice therapy may be length of time spent living full time as a woman before the onset of voice treatment. Even when clients have no formal training in voice feminization, the need to have a voice that does not conflict with a feminine physical appearance provides strong motivation to develop a feminine speaking style. Clients living full time as women would presumably have a lot of trial-and-error experience in attempting feminine voice patterns. This experience might facilitate the progress of a TG client who had lived for years as a woman before seeking voice therapy compared with a client who had not lived full time as a woman.

Such a speculation is somewhat supported by comparing the findings of Gelfer and Tice<sup>9</sup> with the findings of the present study. Gelfer and Tice's subjects had a mean duration of 2 years and 2 months (range = 10 months–4 years, 1 month) living as a woman. Their progress in terms of acoustic and perceptual aspects of voice feminization was superior to the subjects in the present study, where one subject had lived as a woman for 7 months, one had just begun living as a woman, and one was still living as a man. This difference in outcomes occurred despite the fact that the subjects in the two studies were roughly comparable in terms of number of treatment sessions received (16

**TABLE 4.**  
**Interstudy Comparison of Results of Progress Made at the Point of Discharge From Treatment**

Parameters	Dacakis <sup>7</sup>	Meszaros et al <sup>6</sup>	Soderpalm et al <sup>8</sup>	Gelfer and Tice <sup>9</sup>	Present Study
Type of therapy	Symptomatic, with a focus on raising SFF	Symptomatic, with a focus on raising SFF	Symptomatic, with some possible elements of a physiological approach (accent method)	Symptomatic, with a focus on raising SFF	Symptomatic, with a focus on raising SFF, plus a physiological component (VFE)
Number of subjects	10	3	9	5	3
Number of sessions (range)	27 (10–90)	38 (32–76)	17 (3–45)	15.4 (13–16)	12 (12–12)
Speech context	Monologue	Reading passage	Reading passage	Reading passage	Reading passage
Mean SFF pretreatment, Hz (ST#)	125.5 (35)*	150.6 (38)	138.8 (37)	123 (35)	122.0 (35)
Mean SFF posttreatment, Hz (ST#)	168.1 (40)	191.2 (43)	149.3 (38)	194 (43)	177.1 (41)
Mean increase, Hz (ST#)	42.6 (5) <sup>†</sup>	40.6 (5)	10.5 (1)	71 (8)	55 (6)

\* Semitone numbers as standardized by the Acoustical Society of America.<sup>32</sup>

<sup>†</sup> A difference measure obtained by subtracting the pretest from the posttest.

vs 12). Thus, both number of sessions and experience living full time as a woman might be important variables in predicting progress in therapy.

In the past decades, progress had been made in providing an evidence base for voice feminization therapy. Past research has begun to give us insights into the aspects of voice that must be changed for MTF TG individuals to be perceived as female, duration of therapy needed for significant voice change, potential progress following a course of therapy, and maintenance of therapy gains 1–4 years after the termination of therapy, although many questions remain. Results of this study suggest that symptomatic voice treatment plays the most important role in voice feminization, although physiological approaches may be used in a complementary way. Much work remains to be done in the area of voice feminization treatment, but continuing work in this area should help move us closer to our goal of helping MTF TG clients acquire a feminine voice.

## REFERENCES

1. Morgan SW. *Transgender life experiences and expressions: a narrative inquiry into identity recognition and development, bodily experiences, relationships with others, and health care experiences* [doctoral thesis]. Milwaukee, WI: University of Wisconsin–Milwaukee; 2003.
2. Van Kesteren P, Gooren LJ, Megens JA. An epidemiological and demographic study of transsexuals in the Netherlands. *Arch Sex Behav*. 1996; 25:589–600.
3. Wilson P, Sharp C, Carr S. The prevalence of gender dysphoria in Scotland: a primary care study. *Br J Gen Pract*. 1999;49:991–992.
4. Adler R. Transgender/transsexual: an understanding. In: Adler R, Hirsch S, Mordaunt M, eds. *Voice and Communication Therapy for the Transgender/Transsexual Client: A Comprehensive Clinical Guide*. San Diego, CA: Plural Publishing Inc.; 2006:1–39.
5. Van Borsel J, De Cuyper G, Rubens R, Destaerke B. Voice problems in female-to-male transsexuals. *Int J Lang Commun Disord*. 2000;35:427–442.
6. Meszaros K, Vitez L, Szabolcs I, Goth M, Kovacs L, Gorombe Z, Hacki T. Efficacy of conservative voice treatment in male-to-female transsexuals. *Folia Phoniatr Logop*. 2005;57:111–118.
7. Dacakis G. Long-term maintenance of fundamental frequency increases in male-to-female transsexuals. *J Voice*. 2000;14:549–556.
8. Soderpalm E, Larsson A, Almquist S. Evaluation of a consecutive group of transsexual individuals referred for vocal intervention in the west of Sweden. *Logoped Phoniatr Vocol*. 2004;29:18–30.
9. Gelfer MP, Tice RM. Perceptual and acoustic outcomes of voice therapy for male-to-female transgender individuals immediately after therapy and 15 months later. *J Voice*. 2013;27:335–347.
10. Gelfer MP. Voice treatment for the male-to-female transgendered client. *Am J Speech Lang Pathol*. 1999;8:201–208.
11. Carew L, Dacakis G, Oates J. The effectiveness of oral resonance therapy on the perception of femininity of voice in male-to-female transsexuals. *J Voice*. 2007;21:591–603.
12. Andrews M. *Manual of Voice Treatment: Pediatrics Through Geriatrics*. 3rd ed. Clifton Park, NY: Thompson-Delmar Learning; 2006.
13. Spencer L. Speech characteristics of male-to-female transsexuals: a perceptual and acoustic study. *Folia Phoniatr (Basel)*. 1988;40:31–42.
14. Wolfe V, Ratusnik D, Smith F, Northrop G. Intonation and fundamental frequency in male-to-female transsexuals. *J Speech Hear Disord*. 1990;55:43–50.
15. Gelfer MP, Schofield KJ. Comparison of acoustic and perceptual measures of voice in male-to-female transsexuals perceived as female versus those perceived as male. *J Voice*. 2000;14:22–33.
16. Gorham-Rowan M, Morris RM. Aerodynamic analysis of male-to-female transgender voice. *J Voice*. 2006;20:251–262.
17. Holmberg E, Oates J, Dacakis G, Grant C. Phonetograms, aerodynamic measurements, self-evaluations, and auditory perceptual ratings of male-to-female transsexual voice. *J Voice*. 2010;24:511–522.

18. Gunzburger D. Voice adaptation by transsexuals. *Clin Linguist Phon*. 1989; 3:163–172.
19. Gelfer MP, Mikos VA. The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *J Voice*. 2005;19:544–554.
20. Van Borsel J, Janssens J, De Bodt M. Breathiness as a feminine voice characteristic: a perceptual approach. *J Voice*. 2009;23:291–294.
21. Stemple J. *Clinical Voice Pathology: Theory and Management*. Columbus, OH: Merrill; 1984.
22. Sauder C, Roy N, Tanner K, Houtz DR, Smith ME. Vocal function exercises for presbylaryngis: a multidimensional assessment of treatment outcomes. *Ann Otol Rhinol Laryngol*. 2010;119:460–467.
23. Pasa G, Oates J, Dacakis G. The relative effectiveness of vocal hygiene training and vocal function exercises in preventing voice disorders in primary school teachers. *Logoped Phoniatr Vocol*. 2007;32:128–140.
24. Gorman S, Weinrich B, Lee L, Stemple J. Aerodynamic changes as a result of vocal function exercises in elderly men. *Laryngoscope*. 2008;118:1900–1903.
25. Nguyen DD, Kenny DT. Randomized controlled trial of vocal function exercises on muscle tension dysphonia in Vietnamese female teachers. *J Otolaryngol Head Neck Surg*. 2009;38:261–276.
26. McNeill E, Wilson JA, Clark S, Deakin J. Perception of voice in the transgender client. *J Voice*. 2008;22:727–733.
27. American Speech-Language-Hearing Association. Guidelines for Audiological Screening. [Guidelines]; 1997: Available at: [www.asha.org/policy](http://www.asha.org/policy). Accessed February 3, 2010.
28. Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol*. 2009; 18:124–132.
29. Fairbanks G. *Voice and Articulation Drillbook*. 2nd ed. New York, NY: Harper & Row; 1960.
30. Stemple J, Glaze L, Gerdeman B. *Clinical Voice Pathology: Theory and Management*. 2nd ed. Columbus, OH: Merrill; 1995.
31. Acoustical Society of America. *American Standard Acoustical Terminology*. New York, NY: Acoustical Society of America; 1960.
32. Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. *J Acoust Soc Am*. 1995;97:3009–3111.
33. *PASW Statistics (Release 18.0)*. Chicago, IL: IBM Corporation; 2010.
34. Schiavetti N, Metz D. *Evaluating Research in Communicative Disorders*. 5th ed. New York, NY: Prentice Hall; 2006.

## APPENDIX

The following Questionnaire on Vocal Function Exercises was given to the transgender subjects to fill out at the conclusion of therapy. The questionnaire as seen by the subjects is on the left side of the page; the subjects' responses (averaged numerical responses and transcribed comments) are on the right.

## Questionnaire on Vocal Function Exercises

On a scale of 1–5, with 1 being “not at all” and 5 being “very much,” please answer the following questions by circling the appropriate number.

1 = “Not at all” and 5 = “Very much”

Questions	Results
<p>1. How carefully were you able to follow the instructions for the Vocal Function Exercises?</p> <p>1    2    3    4    5</p> <p>Comment:</p> <p>_____</p>	<p>Mean = 5.00</p> <p>No comments</p>
<p>2. How difficult was it for you to do the Vocal Function Exercises?</p> <p>1    2    3    4    5</p> <p>Comment:</p> <p>_____</p>	<p>Mean = 1.67</p> <p>Difficult at first.</p> <p>Did during my daily commute.</p>
<p>3. Did you experience any pain or discomfort when doing the Vocal Function Exercises?</p> <p>1    2    3    4    5</p> <p>Comment:</p> <p>_____</p>	<p>Mean = 3.00</p> <p>A little pain/roughness early on, but it got better.</p>
<p>4. Did you notice any improvement in your breath support or ability to talk longer on a single breath during the time or your participation in the study?</p> <p>1    2    3    4    5</p> <p>Comment:</p> <p>_____</p>	<p>Mean = 4.33</p> <p>Helped me identify breathing issues I have.</p>

<p>5. Did you notice any increase in your speaking pitch during the time of your participation in the study?</p> <p>1    2    3    4    5</p> <p>Comment:</p> <hr/>	<p>Mean = 5.00</p> <p>They helped a lot.</p> <p>I felt I made significant progress.</p>
<p>6. Did you notice that it was harder to speak at a higher pitch during the time of your participation in the study?</p> <p>1    2    3    4    5</p> <p>Comment:</p> <hr/>	<p>Mean = 2.33</p> <p>Not necessarily harder, but it will take practice to be more consistent.</p>
<p>7. Did you find your voice fatiguing more quickly or more often during the time of your participation in the study?</p> <p>1    2    3    4    5</p> <p>Comment:</p> <hr/>	<p>Mean: 1.00</p> <p>My voice got stronger as the study went on.</p>
<p>8. Do you think the Vocal Function Exercises had a positive effect on your use of a feminine voice during the time of your participation in the study?</p> <p>1    2    3    4    5</p> <p>Comment:</p> <hr/>	<p>Mean: 5.00</p> <p>Very positive effect.</p> <p>Yes, it helped improve my vocal awareness.</p>

<p>9. Did you find that the effort it took to do the Vocal Function Exercises daily was balanced by the positive outcome that came from doing them?</p> <p>1      2      3      4      5</p> <p>Comment:</p> <hr/>	<p>Mean: 5.00</p> <p>Yes, very much.</p> <p>At first I didn't see the point, but later on I realized they helped.</p>
<p>10. Do you think the Vocal Function Exercises alone would contribute to the creation and maintenance of a more feminine voice?</p> <p>1      2      3      4      5</p> <p>Comment:</p> <hr/>	<p>Mean: 3.33</p> <p>The exercises helped, but there is no way the same result would have happened without the one-on-one therapy sessions.</p> <p>The exercises need to be combined with lots of practice.</p> <p>I felt the combination of the two approaches was best for me. I feel the more personal feedback I received in the in-person sessions was extremely helpful.</p>